FISHPOND

Comparison of Optimizers in Neural Networks

2016, DEC 07

Abstract

Recent state-of-art successes in deep learning have shown that neural network is a powerful model for various tasks. Since training a deep neural network is usually complex and time-consuming, many methods are applied to speed up this process. As for optimization in neural network, gradient based algorithm is used as the optimizer which coordinates the forward inference and backward gradients of network and reduces the loss. Different gradient algorithms differ at the convergence rate and final performance. This blog gives a survey and some analysis on these algorithms by making a comparison in different datasets with different network architectures.

Introduction

In machine learning applications, gradient based algorithms are popular for training neural networks. Recently, many improved algorithms based on Gradient Descent were presented to boost up the training process like Nesterov, Adagrad, Rmsprop, Adadelta and Adam. All of them are first-order optimization methods and they perform parameter updating during backpropagation to reduce loss of network.

In this work, we first introduce the different variants of Gradient Descent